FUNDAÇÃO GETULIO VARGAS
Escola de Economia de São Paulo
Mestrado em Economia de Empresas

LUCAS JEAN DE MIRANDA COELHO

THE EFFECTS OF TRANSPORTATION INFRASTRUCTURE ON ENTREPRENEURSHIP: EVIDENCE FROM BRAZIL

São Paulo
2025

LUCAS JEAN DE MIRANDA COELHO

THE EFFECTS OF TRANSPORTATION INFRASTRUCTURE ON ENTREPRENEURSHIP: EVIDENCE FROM BRAZIL

Dissertação apresentada à Escola de Economia de São Paulo como requisito à obtenção do título de Mestre em Economia de Empresas.

Área de Concentração: Economia de Empresas
Orientador: Daniel Ferreira Pereira Gonçalves da Mata

São Paulo

2025

Dados Internacionais de Catalogação na Publicação (CIP)
Ficha catalográfica elaborada pelo Sistema de Bibliotecas/FGV

Coelho, Lucas Jean de Miranda.
The effects of transportation infrastructure on entrepreneurship : evidence from Brazil / Lucas Jean de Miranda Coelho. - 2025.
49.f

Orientador: Daniel da Mata.
Dissertação (mestrado CME) – Fundação Getulio Vargas, Escola de Economia de São Paulo.

1. Transportes coletivos. 2. Empreendedorismo. 3. Infraestrutura (Economia). 4. Economia urbana. I. Mata, Daniel da. II. Dissertação (mestrado CME) – Escola de Economia de São Paulo. III. Fundação Getulio Vargas. IV. Título.

CDU 656.121

Ficha Catalográfica elaborada por: Isabele Oliveira dos Santos Garcia CRB SP-010191/O

ERRATA

Autor: Lucas Jean de Miranda Coelho
Título: *The Effects of Transportation Infrastructure on Entrepreneurship: Evidence from Brazil*
Natureza: Dissertação (Mestrado em Economia de Empresas)
Instituição: Fundação Getulio Vargas (FGV) — Escola de Economia de São Paulo (FGV-EESP)
Local e ano: São Paulo, 2025

Finalidade. Esta errata tem por finalidade retificar o texto depositado, exclusivamente na seção *Agradecimentos*, para inclusão do número do processo de fomento, em conformidade com o Termo de Outorga da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

Onde se lê, leia-se conforme indicado:

Folha	Linha(s)	Onde se lê	Leia-se
4	**18-20**	Agradeço também à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), cujo apoio complementar foi fundamental para a conclusão desta dissertação.	Agradeço também à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), **processo nº 2024/10980-7**, cujo apoio complementar foi fundamental para a conclusão desta dissertação.

São Paulo, 12 de janeiro de 2026.

Lucas Jean de Miranda Coelho

LUCAS JEAN DE MIRANDA COELHO

THE EFFECTS OF TRANSPORTATION INFRASTRUCTURE ON ENTREPRENEURSHIP: EVIDENCE FROM BRAZIL

Dissertação apresentada à Escola de Economia de São Paulo como requisito à obtenção do título de Mestre em Economia de Empresas.

Área de Concentração: Economia de Empresas

Banca examinadora:

Prof. Daniel Ferreira Pereira Gonçalves da Mata
FGV-EESP (Orientador)

Prof. Lucas Finamor
FGV-EESP

Prof. Ciro Biderman
FGV-EAESP

AGRADECIMENTOS

Agradeço primeiramente à minha família, em especial à minha mãe Fernanda, à minha avó Beatriz (in memoriam) e à minha tia Nilza, que me acolheu com carinho e apoio quando iniciei o mestrado em São Paulo.

Sou grato aos meus amigos, que estiveram presentes durante esse processo, em especial ao Eliakyn, cuja ajuda foi fundamental ao longo do mestrado.

Minha gratidão também à minha namorada, Fabiana, por estar ao meu lado em todos os momentos, com paciência e apoio incondicional.

Agradeço ao meu orientador, Professor Daniel Ferreira Pereira Gonçalves da Mata, pela orientação dedicada, pelas sugestões valiosas, pela confiança no desenvolvimento deste trabalho e pela generosidade intelectual durante todo o processo.

Estendo meus agradecimentos aos professores da Escola de Economia de São Paulo da Fundação Getulio Vargas, que contribuíram para minha formação com excelência acadêmica e intelectual.

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro por meio da concessão de bolsa de estudos durante parte significativa da pesquisa.

Agradeço também à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), cujo apoio complementar foi fundamental para a conclusão desta dissertação.

Por fim, deixo meu reconhecimento a todos que, direta ou indiretamente, contribuíram para a realização deste trabalho.

RESUMO

.

Este estudo examina o impacto da infraestrutura de transporte no empreendedorismo, analisando a expansão de estações de metrô e trem em São Paulo. Utilizando uma abordagem de Diferenças-em-Diferenças (DiD) com adoção escalonada, estima-se o efeito das novas estações na criação de empresas no nível do setor censitário. Os resultados indicam um aumento significativo no número de empresas próximas às estações recém-inauguradas, com os efeitos mais fortes observados no setor de serviços e entre pequenas empresas. Os resultados sugerem que a melhoria na mobilidade dos trabalhadores, a expansão do mercado consumidor e a aglomeração econômica local são os principais mecanismos impulsionando esse crescimento. O estudo contribui para a economia urbana ao fornecer evidências empíricas sobre como investimentos em transporte público estimulam o desenvolvimento econômico.

Palavras-chave: Transporte público, empreendedorismo, infraestrutura, economia urbana, diferenças-em-diferenças.

ABSTRACT

.

This study examines the impact of transportation infrastructure on entrepreneurship by analyzing the expansion of metro and train stations in São Paulo. Using a staggered difference-in-differences (DiD) approach, we estimate the effect of new stations on firm creation at the census tract level. Results indicate a significant increase in the number of businesses near newly opened stations, with the strongest effects observed in the service sector and small enterprises. The findings suggest that improved mobility for workers, consumer market expansion, and local economic agglomeration are key mechanisms driving this growth. The study contributes to urban economics by providing empirical evidence on how public transportation investments foster economic development.

Palavras-chave: Public transportation, entrepreneurship, infrastructure, urban economics, difference-in-differences.

List of Figures

1 Number of Subway Stations Constructed Over Time 18

2 Relationship between Subway Stations and the Distribution of Establishments 19

3 Active Firms in São Paulo 19

4 Firms Sectors and Firms Sizes 20

5 Distribution of Geocoding Precision for Companies 23

6 Spatial distribution (absolute scale) of the number of firms around Primavera–Interlagos station in 2006 and 2010 26

7 Average Treatment Effects Over Time for Each Length 29

8 Average Treatment Effects by Adoption Year Group 31

9 Distribution of firms at different distances from the Primavera–Interlagos station. 34

List of Tables

1 Descriptive Statistics 21

2 Average Effect of Constructing a Station Across All Treated Census Tracts 28

3 Average Treatment Effects by Period 30

4 Average Treatment Effects by Adoption Group 32

5 Heterogeneity of Treatment Effects Across Sectors 32

6 Heterogeneity of Treatment Effects by Firm Size 33

7 Heterogeneity of Treatment Effects by Distance 35

8 Summary of Treatment Effects by Labor Intensity 37

9 Summary of Treatment Effects 38

10 Summary of Treatment Effects 39

11 Summary of Treatment Effects 40

12 Treatment Effects on Population Density – Mechanism: Population Density 42

Contents

Agradecimentos **i**

1 Introduction **10**

2 Literature Review **13**

3 Institutional Background **16**

3.1 Construction of Metro Stations 16

3.2 Construction of Stations Over Time 17

3.3 Entrepreneurship 17

3.4 Spatial Relation 18

4 Data **20**

4.1 Geocoding Addresses to Census Tract 22

5 Empirical Strategy **23**

6 Results **26**

6.1 Average Effect of Constructing a Station - ATT 28

6.2 Average Effects for Each Length of Exposure 29

6.3 Average Effects for Each Adoption-Year Group 31

6.4 Heterogeneity Analysis 32

7 Mechanisms **36**

7.1 Increase in Labor Mobility and Access to Workforce 36

7.2 Expansion of Access to Consumer Markets 37

7.3 Development of Local Economic Linkages 39

7.4 Population Density 41

8 Robustness **43**

9 Conclusion **44**

1 Introduction

Workers spend approximately 8% of their workday commuting to and from work (Redding and Turner, 2015). Therefore, transportation infrastructure can have a substantial economic impact, as various costs shape the terms of contracts between workers and firms. Railways, for example, have been shown to significantly affect nominal and real economic variables (Donaldson, 2018). However, less is known about how transportation infrastructure influences firm creation, does a new train or subway station spur business growth in surrounding areas?

This study investigates how transportation infrastructure impacts firm creation by analyzing the construction of new subway and train stations in São Paulo. Specifically, we assess how the opening of a station influences entrepreneurship in surrounding census tracts using a Differences-in-Differences (DiD) approach with staggered adoption (Callaway and Sant'Anna, 2021). A key contribution of this research is the incorporation of geocoded data of business openings, with the census sector as the observational unit. Additionally, we examine the underlying mechanisms driving this effect through a more in-depth analysis. Despite growing interest in urban economics, few studies have systematically assessed how transportation infrastructure shapes firm dynamics.

In most countries significant investments are continuing to be made in transportation infrastructure. (Fenske et al., 2023) examine how railway lines in India influence labor market outcomes, providing insights into the relationship between transportation and economic dynamics. One explanation for why transportation shocks influence real and nominal variables is the improved access to regional markets (Donaldson and Hornbeck, 2016). More specifically, market access refers to the extent to which a given locality is exposed to the supply and demand of other regions. This can also benefits production factors as workers, thereby enhancing overall economic efficiency.

When a transportation shock facilitates regional market access, production costs and prices tend to decrease, leading to higher economic growth—a classic positive supply shock (Donaldson and Hornbeck, 2016; Redding and Sturm, 2008).

In São Paulo, subway and train lines facilites the daily movement of millions across the city. According to data from São Paulo's Open Government Companhia do Metropolitano de São Paulo (2025), subway lines alone transported an average of approximately 2.9 million passengers per weekday in 2023—representing around 18% of São Paulo's total population. This means a great concentration of commuters in some places with stations, probably unleashing agglomeration economies.

According to Companhia do Metropolitano de São Paulo (2023), the metro accounted for 22.6% of all trips made using public transport modes in the São Paulo Metropolitan Region (RMSP) in 2023, slightly higher than the 22.2% recorded in 2017. Furthermore, the average commuting time using public transportation—which includes metro, trains, and buses—was approximately 58 minutes per leg, suggesting a total of around 1 hour and 56 minutes spent daily on round trips. These figures underscore the central role of rail infrastructure in shaping urban mobility and highlight the economic relevance of evaluating the impacts of such investments in the São Paulo metropolitan area.

The results show that the opening of metro and train stations in São Paulo had a positive and significant impact on firm creation in nearby census tracts. This effect is particularly strong among small businesses and in the service sector. The analysis suggests that improved worker mobility, expanded access to local consumer markets, and intensified economic agglomeration are the main mechanisms driving this growth. These results suggest that investments in public transportation infrastructure may help stimulate entrepreneurship in nearby areas, particularly in underdeveloped or underserved urban regions.

According to data from Data Sebrae - Serviço Brasileiro de Apoio às Micro e Pequenas Empresas (SEBRAE) (2025), Brazil's business landscape is predominantly composed of small and medium-sized enterprises, with individual microentrepreneurs (MEIs) and microenterprises (MEs) accounting for 85.28% of all businesses. A significant share of these enterprises are highly labor intensive, often operating as individual ventures or employing only a small number of workers. In particular, 82% of MEI and ME are concentrated in the commerce and services sector, strengthening their pivotal role in urban growth and economic development.

Understanding the expansion of these businesses to analyze the broader dynamics of urbanization and, simultaneously, the extent to which transportation influences economic activity.

In short, improved transportation infrastructure leads to development due to four factors. Firstly, there's a reduction in transaction costs and consequent market integration. Secondly, there's a promotion of mobility of production factors, including labor. Thirdly, human capital is better utilized, meaning you can access better health and education opportunities for your children. You can work in Neighborhood A, send your children to Neighborhood B to study, and live in Neighborhood C, using the potential of each region. Lastly, there are the benefits of agglomeration, such as increased information flow and generation of new ideas (Banerjee et al., 2014; Duranton and Puga, 2020).

The focus of this study is to analyze, using the difference-in-differences approach with staggered adoption (Callaway and Sant'Anna, 2021), how the opening of a specific subway station influences entrepreneurship in the surrounding area. This allows us to estimate both overall and dynamic causal effects across groups and over time. This methodology is particularly suited to our context, as the gradual opening of subway stations across census tracts provides plausibly exogenous variation, which we exploit using a quasi-experimental approach."

The study finds that the opening of new metro and train stations in São Paulo led to an average increase of approximately 39 firms in nearby census tracts. This result reflects a substantial absolute effect, especially considering the typically low baseline number of firms in those areas prior to the treatment. The effects are strongest among small businesses and in the service sector, with significant impacts concentrated within 500 meters of the stations. Moreover, the positive effects become statistically significant starting in the fourth year after a station opens, suggesting a delayed response in firm creation. Mechanisms such as improved labor mobility, increased population density, and expanded access to consumer markets help explain these outcomes.

This paper contributes to the literature by identifying and measuring a previously unexplored effect of transportation infrastructure on new firm formation. Addition-

ally, we conduct heterogeneity analyses to examine how this impact varies across different contexts. Our analysis draws on insights from spatial equilibrium models, agglomeration effects, and firm dynamics to offer a novel empirical perspective on how transportation investments can boost entrepreneurship.

The remainder of this thesis is structured as follows. Section 2 reviews the relevant literature on transportation infrastructure, entrepreneurship, and urban economics. Section 3 provides the institutional background, including an overview of station construction and the evolution of subway infrastructure in São Paulo. Section 4 describes the data sources and geocoding procedures used to match firms to census tracts. Section 5 outlines the empirical strategy based on the staggered difference-in-differences approach. Section 6 presents the main results, followed by Section 7, which explores potential mechanisms underlying the observed effects. Section 8 discusses robustness checks, and Section 9 concludes the study.

2 Literature Review

This study contributes to the literature on spatial economic equilibrium models, agglomeration effects, and the impacts of infrastructure on economic growth, with a specific focus on firm creation. Although much research has explored how transportation infrastructure affects productivity, employment, and population dynamics, few empirical studies have examined its direct impact on entrepreneurship, which we define as firm openings.

Economic geography plays a crucial role in the distribution of economic activity, influencing the location and growth of firms. Allen and Arkolakis (2014) show that at least 20% of a region's economic activity can be attributed to geographical factors, with investments in transportation infrastructure, such as highways, generating welfare gains that exceed their costs by improving labor mobility and regional connectivity.

Desmet and Rossi-Hansberg (2014) highlight that transportation costs and technology diffusion shape sectoral concentration and productivity. These changes can lead to lasting economic transformations in specific regions.

Agglomeration is also important for economic growth. Felkner and Townsend (2011) demonstrate that firm concentration predicts future growth, while Gaubert (2018) show that geography influences firm productivity through selection effects and agglomeration economies.

Despite these insights, few studies have focused on the impact of transportation infrastructure on firm creation. Research by Donaldson and Hornbeck (2016) and Fenske et al. (2023) shows that access to railways and transit improves market connectivity but does not directly address new business formation. Similarly, entrepreneurship studies such as Beaudry et al. (2018) and Glaeser et al. (2015) do not isolate the causal effect of transportation infrastructure.

Promoting entrepreneurship requires policies that directly engage with urban structure. Faggio and Overman (2014) argue that transportation policies can act as a "big push", expanding access to opportunities and reinforcing regional benefits.

Fenske et al. (2023) and Donaldson and Hornbeck (2016) show that transportation infrastructure expands market access and boosts population growth, which can drive economic growth.

Haddad et al. (2015) analyze the benefits of train and metro stations on labor productivity at the municipal level, demonstrating that removing these infrastructures would significantly impact output, especially in metropolitan regions like São Paulo, where peripheral cities rely on central urban hubs.

Transportation policies can serve as instruments for local economic growth by reducing distance and travel time, directly affecting prices (Ahn et al., 2020), increasing transaction volumes, and facilitating business expansion. Melo et al. (2009) argue that proximity between workers and workplaces increases productivity, fostering firm creation. Less dense regions face structural challenges, but mobility can overcome these barriers, enhancing interactions between regions (Prud'Homme and Bocarejo, 2005).

Amior and Manning (2019) highlight that transportation policies reduce labor market barriers, influencing job location decisions. Beaudry et al. (2018) show that firm formation increases with population density and employment growth, though wage pressures initially moderate this effect.

Studies like Glaeser et al. (2015) emphasize that business clusters drive innovation and productivity, shaping effective public policies. Faggio and Overman (2014) contrast self-employment in rural areas with entrepreneurial dynamism in urban centers, underscoring the importance of local interactions.

Beyond classical models of agglomeration economies and market access, more recent literature has emphasized the importance of micro-level mechanisms in shaping the economic impact of urban infrastructure. Jacobs (1961), for example, argues that the diversification of land use and the presence of people on the streets foster not only economic vibrancy but also urban safety and innovation. These arguments suggest that the spatial distribution and diversification of businesses around transit hubs may themselves generate dynamic spillovers, a hypothesis particularly relevant to dense urban areas like São Paulo.

Complementary empirical evidence supports this granular perspective. Atkin et al. (2018), studying the diffusion of mobile phone networks, show that localized infrastructure shocks can alter consumer behavior, firm entry, and market structure even in narrowly defined geographies. These studies underscore that spatially disaggregated analyses, such as those using census tracts, are not only methodologically viable but essential to uncover heterogeneous treatment effects at the local level.

Evaluating the impact of public infrastructure investments, such as station construction, is challenging due to unobserved factors and selection biases. This study employs a staggered difference-in-differences methodology to generate causal estimates, testing assumptions to ensure validity.

This paper fills a gap in the literature by providing causal evidence on how investments in transportation infrastructure affect firm creation at a granular spatial level. While prior research has examined transportation's effects on productivity, wages, and population, few have focused on its role in fostering entrepreneurship. By combining geocoded firm-level data with a staggered DiD framework, this study isolates the effect of new transit stations on business formation in their immediate surroundings. To the best of our knowledge, no previous work has investigated the effects of transportation infrastructure on firm creation.

3 Institutional Background

3.1 Construction of Metro Stations

The construction of a metro station is a process that unfolds in multiple stages, from initial planning to full operation. The timeline for each phase can vary significantly, influenced by project complexity, local conditions, and specific challenges. Below, we provide an overview of the key phases along with their estimated durations, based on relevant sources.

The initial planning and feasibility studies phase includes identifying project needs, conducting demand analyses, assessing environmental impacts, and evaluating economic feasibility, with its duration varying based on project complexity and administrative efficiency (Vuchic, 2005). Next, the licensing and approvals stage involves securing environmental permits, obtaining government clearances, and conducting public consultations (Flyvbjerg, 2003). Finally, the financing and bidding process entails securing funding and selecting construction firms through bidding procedures, with its duration depending on the chosen financing model and the length of the procurement process (World Bank, 2019).

The construction phase is typically the longest among all stages—longer than both the planning and post-construction periods—and its duration can vary considerably depending on project scale, construction methods, and geological conditions. The average construction time for a metro station depends on the structure type and technology applied:

- **Underground stations (TBM or NATM):** 3 to 5 years per station (International Tunneling and Underground Space Association, 2020).
- **Elevated stations:** 2 to 4 years (World Bank, 2018).
- **Surface-level stations:** 1.5 to 3 years (Albalate and Bel, 2012).[1]

For example, the construction of São Paulo Metro's Line 4-Yellow took an average of **4 years per station** (Couto, 2011). The extension of Line 2-Green (Vila

[1] We are gonna use this to check robustness

Prudente – Penha) is expected to take **3 to 4 years per station** (Companhia do Metropolitano de São Paulo, 2022). Metro expansions in London and New York indicate average timelines between **3 and 6 years** for each deep underground station (Flyvbjerg and Gardner, 2023).

The testing and initial operation phase begins after construction, involving operational tests and personnel training to guarantee safety and efficiency before public opening (International Association of Public Transport, 2018).[2]

It is important to note that the exact timing of station completion is often subject to political shifts, administrative hurdles, financial constraints, and engineering challenges. These factors introduce substantial uncertainty into when a station becomes operational. Consequently, the date of station opening can be considered plausibly exogenous to local firm-level decisions or entrepreneurial activity. This characteristic strengthens our empirical identification, as we leverage the variation in station opening dates to estimate causal impacts on firm creation.

3.2 Construction of Stations Over Time

The construction of subway stations in São Paulo provides a natural setting to analyze the effects of transportation infrastructure on firm creation. For data quality reasons, the most suitable period for analysis begins in 1993. Figure 1 shows the evolution of subway station openings over time.

This temporal variation in station construction allows us to leverage a staggered adoption Differences-in-Differences (DiD) approach to estimate the causal effects of new subway stations on firm creation in nearby census tracts.

3.3 Entrepreneurship

Entrepreneurship refers to the process of identifying, developing, and capitalizing on opportunities to create new businesses, products, or services, involving risk-taking and innovation to generate economic and social value (Shane et al., 2003).

[2]Various factors can impact project timelines, including scope changes, technical challenges, environmental issues, and political shifts. Additionally, delays are frequent in large infrastructure projects, often exceeding initial estimates (Flyvbjerg, 2009).

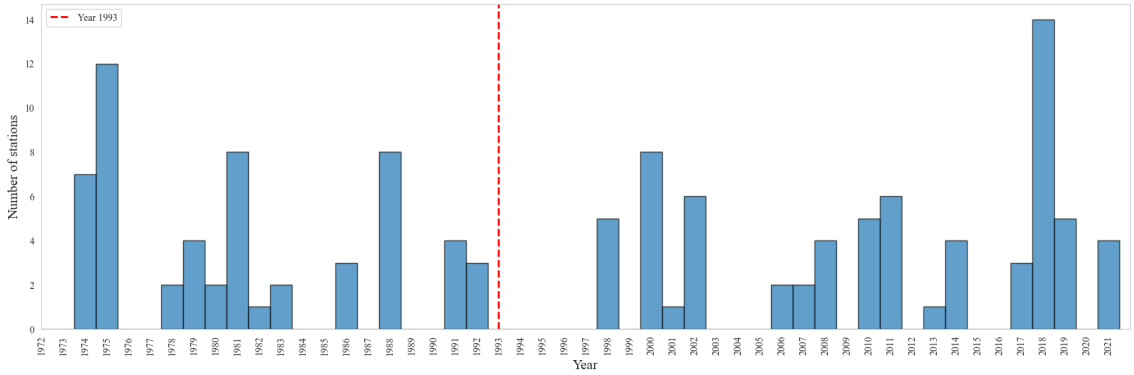

Figure 1: Number of Subway Stations Constructed Over Time

Notes: This figure presents the cumulative number of subway stations constructed over time. The trend provides insights into the expansion of the transit network and its potential economic impact. Years with rapid increases indicate periods of significant infrastructure investment.

According to Drucker (1985), entrepreneurship extends beyond starting new companies, encompassing innovation within existing organizations, a concept known as intrapreneurship.

> "The entrepreneur always searches for change, responds to it, and exploits it as an opportunity." (Drucker, 1985, p. 25)

Here, we adopt the basic definition of entrepreneurship as the creation of new firms. While simple, this definition enables the application of various econometric methods to effectively measure entrepreneurship.

3.4 Spatial Relation

To investigate the connection between subway stations and firm distribution, we assess the spatial distribution of establishments in relation to station locations. Figure 2 visually represents this relationship, demonstrating how businesses concentrate around subway stations.

The figure reveals a clear concentration of firms around subway stations, indicating that proximity to transportation infrastructure may contribute to economic activity. Areas with greater accessibility may attract more establishments due to enhanced mobility for workers, customers, and suppliers. This pattern aligns with urban economic theories, which emphasize the role of transportation infrastructure

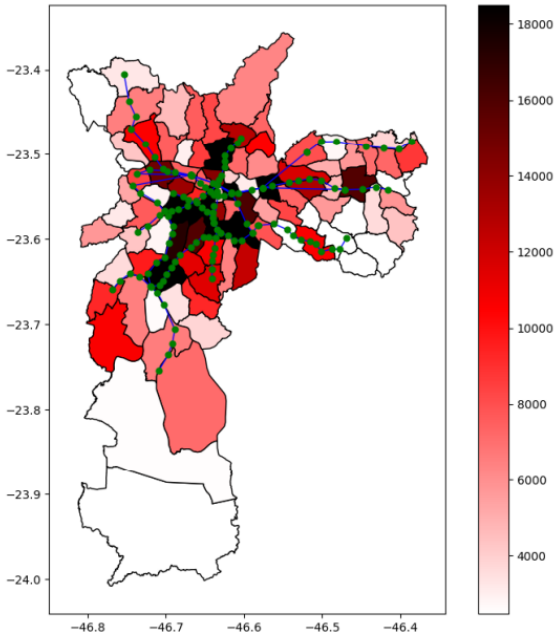

Figure 2: Relationship between Subway Stations and the Distribution of Establishments

Notes: This figure illustrates the spatial distribution of establishments relative to the location of subway stations. Darker areas indicate a higher concentration of businesses. The visualization helps to assess the potential economic impact of transit accessibility.

in shaping economic geography, highlighting the need for rigorous econometric analysis to test and quantify its impact.

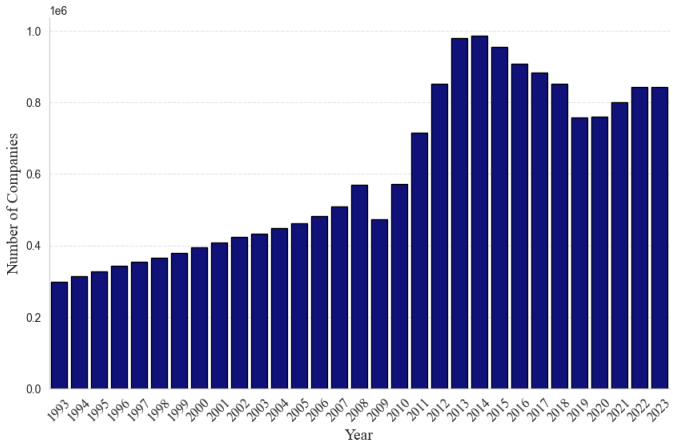

Figure 3: Active Firms in São Paulo

Notes: This graph presents the evolution of the number of companies over the years. Growth or decline may be associated with macroeconomic factors, public policies, or structural changes in the economy. The Y coordinate (number of companies) is a scale of 1 million.

Figure 3 highlights the growth and decline patterns in the number of firms, which

may be more influenced by macroeconomic factor or structural changes. Specifically, the break in the growth rate of firms in 2010 is linked to the introduction of MEI companies, which also served as an alternative for businesses that preferred not to hire employees formally. Whereas the further declines in the number of firms coincide with economic recessions in the country (starting at 2014).

(a) (b)

Figure 4: Firms Sectors and Firms Sizes

Notes: These figures illustrate key aspects of the productive structure and economic dynamics of the analyzed regions.

Figure 4 shows that services and commerce dominate, while industry and unspecified sectors have lower representation, possibly due to a smaller industrial base or missing data. Figure 5 reveals that small firms prevail, followed by medium-sized firms, with large firms being the least common. Together, these trends indicate that the analyzed region (São Paulo) are driven by local enterprises, which shapes economic dynamics, employment, and transportation needs.

4 Data

This study integrates multiple data sources to examine the impact of transportation infrastructure on firm creation.

Entrepreneurship Data: The publicly available business registration data from the Federal Revenue provides detailed information about formal firms operating in Brazil. We performed annual extractions of this dataset for each year from 2000 to 2022, using the public CNPJ registry available through the Federal Revenue's open data platform. The database includes the start date of firm activities, company size,

registered capital, CNAE (economic activity code), full address, contact information, and the date of any record modification. For closed firms, this date corresponds to the closing date, and a dedicated column specifies the reason for the change.

Subway and Train Stations Data: Information on subway and train stations was gathered from two primary sources: the transparency portal of the São Paulo Subway system and the "Our History" portal of CPTM (Companhia Paulista de Trens Metropolitanos).

Control Variables Data: To mitigate potential confounding factors, we include control variables aggregated at the census tract level. These are sourced from the 1991, 2000, 2010, and 2022 Population Censuses conducted by the Brazilian Institute of Geography and Statistics (IBGE).[3]

Table 1: Descriptive Statistics

Variable	**Mean**	**Std. Dev.**	**Min**	**Median**	**Max**
area	73,371.39	446,034.85	453.39	37,966.85	47,011,590.13
population	690.38	314.03	0.00	658.30	7,176.00
households	225.28	94.74	0.00	221.67	7,175.00
income	324,407.33	421,647.27	0.00	200,652.00	6,877,215.00
companies	37.85	99.29	1.00	21.00	12,872.00

Notes: This table presents descriptive statistics for key variables, aggregated at the census tract level. *area* is measured in square meters; *population* is the total number of residents; *households* is the number of households; *income* is the total household income in current BRL; and *companies* is the number of active firms per census tract. For instance, the median income per census tract is BRL 200,652, reflecting the high income heterogeneity across São Paulo's neighborhoods.

Table 1 presents summary statistics for key variables. Overall, the data indicate substantial spatial and economic heterogeneity, which should be accounted for in further analysis. [4]

[3]We applied linear interpolation to estimate values for intercensal years. While this method assumes a constant rate of change over time, it is a standard approach given the decennial nature of census data.

[4]They are will be our control variables

4.1 Geocoding Addresses to Census Tract

First, we performed data collection by obtaining address data from the Federal Revenue. Next, data cleaning ensured consistency across records by standardizing address formats. The geocoding process was then conducted using the "geocodebr" package, a recent addition to the R software ecosystem.[5] This process provided geographic coordinates (latitude and longitude) for each address. Following this, mapping to census sectors was carried out using census tract boundary files in shapefile format (.shp), where spatial joins (point-to-polygon method) were applied to assign addresses to their respective census tracts.

[5]The package relies on open spatial datasets of Brazilian addresses, primarily using the Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE), which is published by the Instituto Brasileiro de Geografia e Estatística (IBGE), Brazil's official agency for statistics and geography.

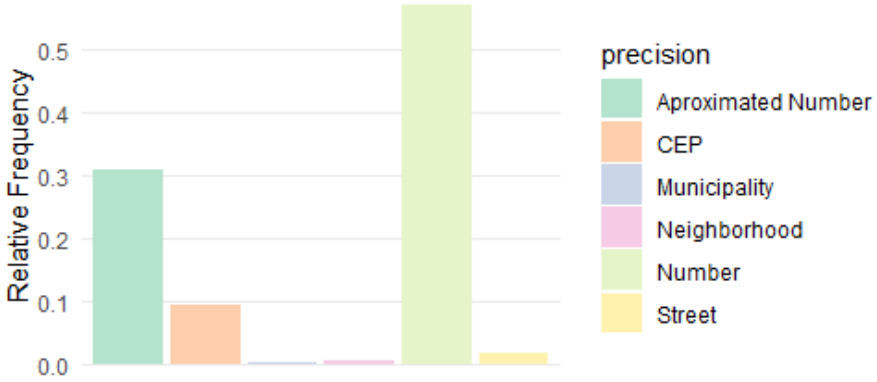

Figure 5: Distribution of Geocoding Precision for Companies

Notes: CEP refers to postal code-level geocoding (It's equivalent to ZIP postal code in the US); *Neighborhood* refers to neighborhood-level geocoding; *Street* refers to street-level geocoding; *Municipality* refers to municipality-level geocoding; *Number* refers to exact address-level geocoding; *Aproximated Number* refers to approximate address-level geocoding for companies.

Figure 6 illustrates the distribution of geocoding precision for companies, highlighting the relative frequency of different accuracy levels. Most addresses were geocoded with high precision, falling under *number* (exact address) and *approximated_number*, which together account for 87,75%. Less precise categories, such as *street*, *neighbourhood*, and *municipality* together represent only 2,62%, appear less frequently, while *cep* (postal code) has moderate representation, 9,6%. This suggests that geocoding was largely effective, though a minority of cases were resolved at broader spatial levels. Each firm was then matched to a specific census tract. These tracts serve as the unit of observation for the analysis, where treatment is defined as the construction and opening of a subway or train station within the tract's boundaries.

5 Empirical Strategy

This study estimates the effect of subway and train station construction on firm creation in São Paulo, using the Callaway and Sant'Anna (2021) Differences-in-Differences (DiD) approach. The treatment is defined as the construction and opening of a subway or train station within the boundaries of a given census tract. A

tract is considered treated from the year the station begins operation. The key features of the treatment are: it is irreversible and staggered, meaning that each census tract receives a station at different points in time, and once constructed, the station remains in place permanently. The treatment is also group-based, as census tracts are assigned to a group G based on their year of adoption.

The key features of the treatment are: the treatment is irreversible and staggered, meaning that each census tract receives a station at different points in time, and once constructed, the station remains in place permanently. The treatment is also group-based, as census tracts are assigned to a group G based on their year of adoption. The control group consists of census tracts not yet treated at time t, including those that are treated in future years or never treated during the period of study.

For each census tract, the potential outcomes are defined as follows: $Y_t(0)$ represents the outcome if the census tract has not yet been treated, while $Y_t(g)$ denotes the outcome if the census tract received a station in year g.

The average treatment effect on the treated (ATT) is then given by:

$$ATT(g,t) = E[Y_t(g) - Y_t(0) \mid G = g]$$

where g represents the year of station construction, and t refers to the calendar year. The term $G = g$ denotes the group of census tracts that received treatment in year g. At each point in time, treated census tracts are compared to those that have not yet received treatment, serving as valid control units under the conditional parallel trends assumption. This staggered setup allows us to leverage variation in treatment timing to identify dynamic effects, comparing outcomes before and after treatment across cohorts.

To capture how the effects evolve over time, we aggregate the treatment effect based on time elapsed since station construction, given by:

$$\theta_e(e) = \sum_g 1\{g + e \leq T\} \cdot P(G = g) \cdot ATT(g, g + e)$$

where $e = t - g$ represents the years since station construction, and $P(G = g)$

denotes the probability of a census tract belonging to group $G = g$.

To assess differences based on when the station was built, we compute:

$$\theta_{sel}(g) = \frac{1}{T - g + 1} \sum_{t=g}^{T} ATT(g, t)$$

where T is the final year of the study, t corresponds to the calendar year, and g represents the year of station construction.

Finally, the overall treatment effect is given by:

$$\theta_O = \sum_g \theta_{sel}(g) \cdot P(G = g)$$

where $\theta_{sel}(g)$ represents the average treatment effect for census tracts treated in year g, and $P(G = g)$ denotes the probability of a census tract receiving treatment in year g. To estimate the group-time ATT, we employ the doubly robust method, which combines outcome regression and inverse probability weighting, providing consistent estimates even if one of the models is misspecified. The analysis covers the period from 1998 to 2023, including all years in which new metro or train stations were inaugurated and corresponding firm data is available. Standard errors follow the robust variance estimator implemented in the `did` package (Callaway and Sant'Anna, 2021), accounting for clustering at the census tract level.

The identification strategy relies on several key assumptions. First, the conditional parallel trends assumption posits that, conditional on control variables, treated and "not-yet-treated" groups would have exhibited similar outcome trends in the absence of treatment. Second, we assume no spillover effects—that is, the construction of a station in one census tract should not affect firm outcomes in untreated tracts, which serve as controls until their own treatment. Third, the no anticipation assumption requires that the treatment has no effect before a station is actually built; if anticipation exists, it must be explicitly modeled. Lastly, the model allows for both homogeneous and heterogeneous treatment effects, meaning the impact of a station may be constant across treated areas or vary by adoption timing and location.

6 Results

To motivate the analysis, Figure 6 presents a descriptive analysis of firm density within census tracts located within a 250-meter radius of the *Primavera–Interlagos* station, considering the years 2006 and 2010 (2 years before and 2 years after the station). An expansion in business presence can be observed around the station, especially after the implementation of Line 9–Emerald, which improved the region's accessibility. Red areas indicate a higher density of firms, while blue markers represent the exact location of the station under analysis.

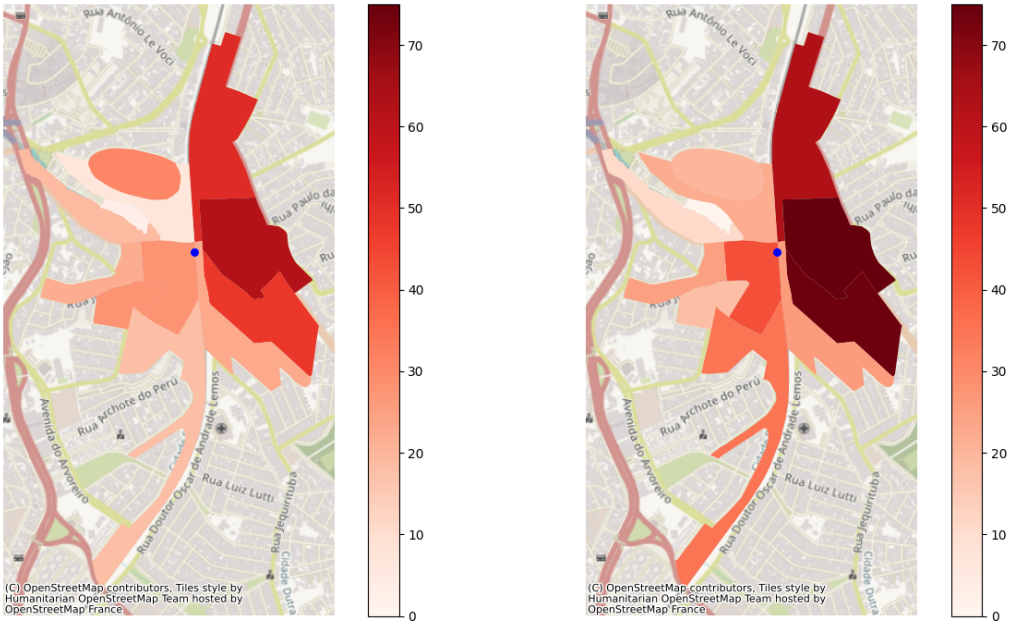

Figure 6: Spatial distribution (absolute scale) of the number of firms around Primavera–Interlagos station in 2006 and 2010

Note: Figure 6 shows the spatial evolution of the number of firms in the vicinity of **Primavera–Interlagos** station. The two panels compare the years 2006 (left) and 2010 (right), using absolute values. Red shading reflects higher business density, and the station is marked with a blue dot.

There is a notable increase in firm density in 2010, particularly in tracts to the northeast and southwest of the station, where deeper red tones emerged. Tracts that previously had fewer firms in 2006 show an increase in business presence in 2010, indicating a growth in economic activity in the area.

This pattern may suggest positive effects from the improved accessibility offered by

the station, such as greater foot traffic, enhanced commercial appeal, and increased land value. Furthermore, the tracts immediately adjacent to the station, which already had a significant number of firms in 2006, maintained or increased their density in 2010, indicating a consolidation of business activity near the infrastructure.

6.1 Average Effect of Constructing a Station - ATT

The table below presents the average effect of constructing a station across all treated units.

Table 2: Average Effect of Constructing a Station Across All Treated Census Tracts

Statistic	**Number of Establishments**
Treatment effect	38.7621**
	(17.2698)
Treated units	40
Groups	15
Observations	8,703
Control Group Mean	49.39952

Notes: The table reports the average treatment effect on the treated (ATT) for census tracts that received a station. The standard error is reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

The findings reveal that the construction of a station significantly increased the number of establishments in the treated census tracts. The estimated average treatment effect on the treated (ATT) is 38,76, indicating that, on average, areas receiving a station experienced an increase of approximately 39 establishments, an increase of approximately 39 establishments, which corresponds to a relative effect size of 0.78 compared to the control group mean. Although the estimated 78% increase appears large, it is important to note that treated tracts initially had a low baseline level of firm activity, often close to zero. Hence, part of this effect reflects the extensive margin of firm creation in previously underserved areas.

These results suggest that enhanced accessibility and infrastructure from the station's construction contributed to increased business activity in the affected areas.

6.2 Average Effects for Each Length of Exposure

The figure below shows the average treatment effects for each length of exposure.

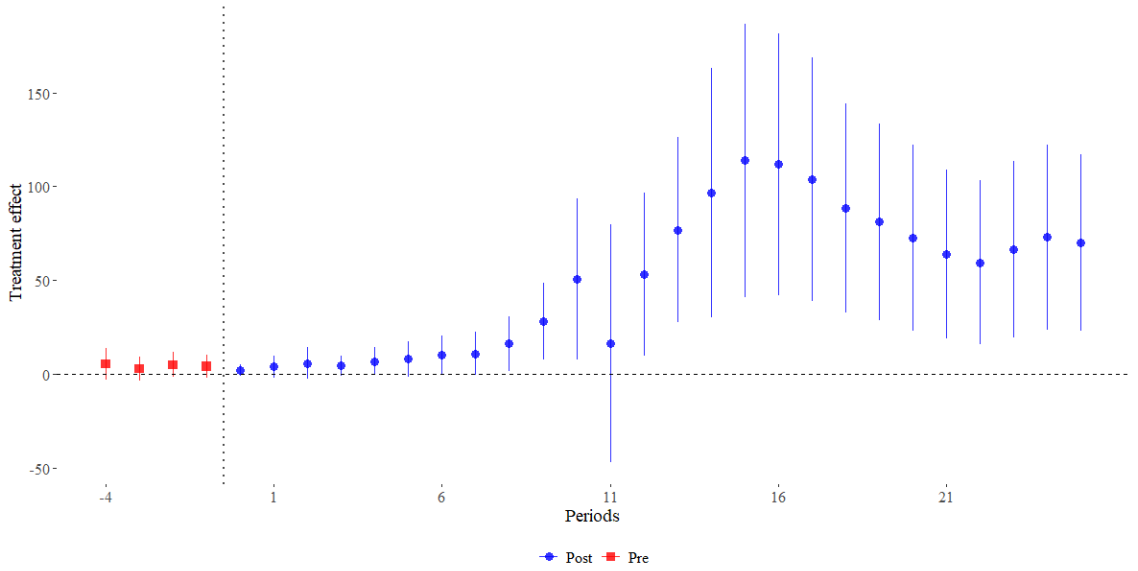

Figure 7: Average Treatment Effects Over Time for Each Length

Notes: This figure illustrates the estimated treatment effects over time for different lengths of exposure. The estimates account for potential heterogeneous effects depending on the duration since treatment implementation. The pre-treatment phase ("Pre" - red) represents time intervals before the intervention, while the post-treatment phase ("Post" - blue) includes periods following the intervention.

This approach allows us to assess the parallel trends assumption as well as the temporal structure of the treatment effects. The estimates show no statistically significant pre-treatment effects, lending support to the identification strategy. Post-treatment, the effects on firm creation appear with a delay and grow in magnitude over time, consistent with the theoretical mechanism of gradual ecosystem formation.

During the pre-treatment period, the ATT values remain relatively small and largely statistically insignificant, indicating no strong pre-existing trends before treatment adoption. However, after the intervention (from period 0 onward), the ATT values begin to rise, with statistically significant effects emerging from period 4 onward. It's visible tha a station can keep it's effect to create new for many years.

The results suggest a growing treatment effect over time, with particularly large and highly significant effects observed in later periods (e.g., period 10 and beyond).

Examining the dataset most effect onward 10 years are estimative since fewer stations in the data was constructed so many time ago. The interesting here is the delay to the effect appears to us. A station can sustain its impact for many years, but why does the effect only emerge after approximately 3 to 4 years? What explains this delay?

Period	ATT	Standard Error	Group
-4	5.48	3.94	Pre
-3	2.78	2.97	Pre
-2	4.94	3.26	Pre
-1	4.21	3.11	Pre
0	2.02	1.70	Post
1	4.01	2.55	Post
2	5.78	4.04	Post
3	4.40	2.75	Post
4	6.80*	3.48	Post
5	8.00*	4.22	Post
6	10.00**	4.85	Post
7	10.90**	5.03	Post
8	16.30***	6.22	Post
9	28.20***	9.39	Post
10	50.60***	22.20	Post
11	16.50	32.60	Post
12	53.10***	20.60	Post
13	76.80***	24.40	Post
14	96.60***	31.50	Post
15	114.00***	35.50	Post
16	112.00***	33.80	Post
17	104.00***	31.20	Post
18	88.50***	27.90	Post
19	81.20***	26.20	Post
20	72.60***	24.90	Post
21	63.80***	22.90	Post
22	59.50***	21.50	Post
23	66.30***	22.00	Post
24	72.90***	24.60	Post
25	70.10***	24.30	Post

Table 3: Average Treatment Effects by Period

Notes: This table presents the estimated average treatment effects (ATT) over different periods relative to treatment adoption. Pre-treatment periods are labeled as "Pre", and post-treatment periods as "Post". Standard errors are reported in the second column. Statistical significance is indicated as follows: *p<0.10, **p<0.05, ***p<0.01.

6.3 Average Effects for Each Adoption-Year Group

This graph shows the treatment effects by year of adoption.

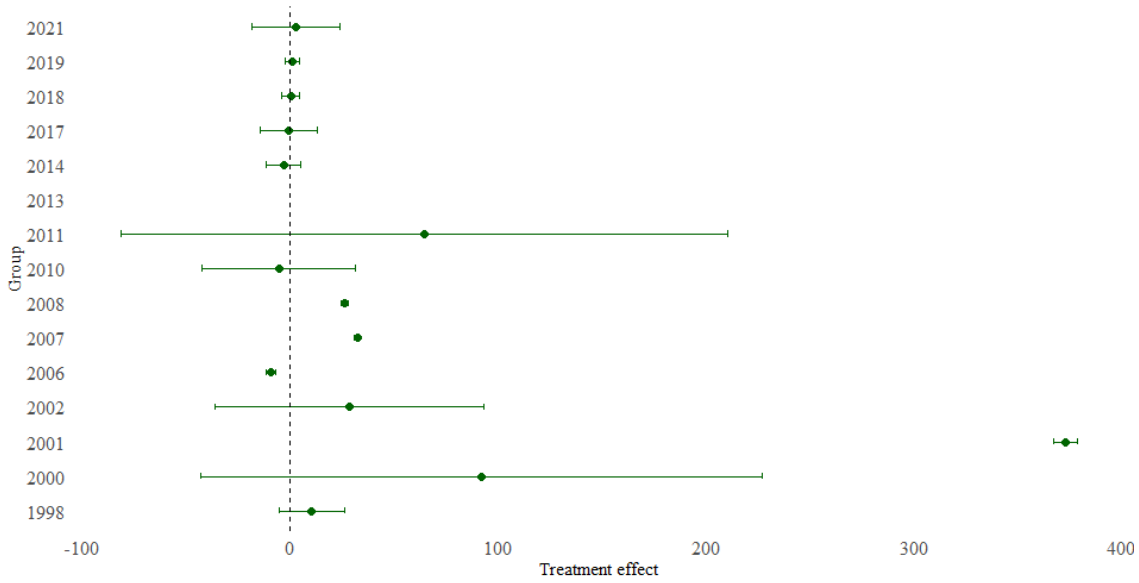

Figure 8: Average Treatment Effects by Adoption Year Group

Notes: This figure presents the estimated average treatment effects (ATT) for different adoption year groups. The estimates capture variations in treatment effects based on the year in which a census tract received the intervention. Confidence intervals at 5% signifcance reflect statistical uncertainty.

The results reveal considerable variation in treatment effects across groups, with some showing positive and statistically significant impacts. Overall, the distribution skews to the right, though not all effects are significant. The 2001, 2007, and 2008 groups display particularly strong and significant increases in firm creation near newly opened stations. However, the 2001 group stands out as an outlier, likely driven by an isolated startup boom in the Vila Olímpia neighborhood — a highly entrepreneurial area that grew rapidly around the turn of the century, coinciding with the opening of a station bearing the same name.

Conversely, certain groups present negative effects, such as the 2006 group (-9.09***), suggesting a potential decline in the analyzed outcome following treatment adoption. This gave us some evidence that there is some heterogeneous factors, the delay is likely influenced by geographic factors, given the relatively high standard errors. This warrants further investigation to inform policy implementation. In relation to the adoption year, the treatment effects appear to be more positive for

earlier years, although the evidence is less conclusive.

Group	ATT	Standard Error
1998	10.3*	5.54
2000	91.8*	64.6
2001	373.00***	3.00
2002	28.4	41.0
2006	-9.09***	1.24
2007	32.4***	0.716
2008	26.2***	0.864
2010	-5.46	21.2
2011	64.6	74.6
2013	–	–
2014	-3.07**	1.36
2017	-0.52	0.577
2018	0.26	2.31
2019	0.94	3.01
2021	2.76	5.57

Table 4: Average Treatment Effects by Adoption Group

Notes: The coefficients represent the average treatment effect for each adoption group. Significance levels are indicated by * p<0.10, ** p<0.05, *** p<0.01. The values "–" indicate the absence of available observations for the corresponding group.

6.4 Heterogeneity Analysis

In this section, we analyze the heterogeneity of the treatment effect across different economic sectors: **Commerce, Industry, and Services**. Table 5 presents the estimated treatment effects (ATT) along with their standard errors.

Table 5: Heterogeneity of Treatment Effects Across Sectors

Measure	Commerce	Industry	Services
ATT	6.3721*	8.8253*	26.5364*
	(3.7036)	(4.8805)	(13.9587)
Treated Units	39	34	38
Groups	15	15	15
Observations	7,549	5,019	7,598
Control Mean	19.2366	8.5954	26.6531

Notes: The ATT estimates represent the average treatment effect on the treated for each economic sector. Standard errors are in parentheses. The number of treated units refers to the census tracts that received treatment. The control mean represents the average number of establishments in the control group. * p<0.10, ** p<0.05, *** p<0.01.

First, the results indicate that for 10% of significance level, a new station effects every sector of firms. But also indicate that the effects of station construction vary significantly across economic sectors. While the impact on **Commerce** and **Industry** is relatively modest, the effect on **Services** is larger, suggesting that stations may have a stronger influence on service-oriented businesses. But, if we consider the level of industry (control mean) the increase is about 102,67%, a very large increase, the same for Services with 99,56%, another expressive increase.

The higher ATT for services, 26.54 firms, suggests that this sector benefits the most from improved accessibility and increased foot traffic generated by the station. On the other hand, the lower ATT for commerce and industry indicates that the presence of a station alone may not be a strong determinant of business growth in these sectors.

These findings align with previous studies on urban infrastructure, which suggest that services tend to benefit more from transit improvements due to increased customer flow and higher flexibility in location choice (Melo et al., 2009; Glaeser and Gottlieb, 2009; Duranton and Puga, 2020).

Table 6: Heterogeneity of Treatment Effects by Firm Size

Measure	**Small Firms**	**Medium Firms**	**Large Firms**
ATT	18.0286**	4.5579**	16.5868
	(8.3164)	(1.9724)	(15.2842)
Treated Units	40	26	37
Groups	15	13	15
Observations	7,830	2,588	6,937
Control Mean	31.0389	7.8429	20.9167

Notes: The ATT estimates represent the average treatment effect on the treated for firms of different sizes. Standard errors are in parentheses. The number of treated units refers to the census tracts that received treatment. The control mean represents the average number of establishments in the control group. * p<0.10, ** p<0.05, *** p<0.01.

The results from Table 6 suggest that small firms experienced the largest and statistically significant effect of the treatment, with an ATT of 18.03 (p<0.05). This suggests that the intervention had a more pronounced impact on smaller firms in terms of establishment growth. Medium-sized firms also exhibited a positive and

statistically significant effect, but the magnitude of the impact was lower (ATT = 4.56, p<0.05). For large firms, the ATT estimate is 16.59, but it is not statistically significant, implying that the observed effect could be due to random variation.

Overall, these findings indicate that the impact of the intervention varies by firm size, with small and medium firms experiencing stronger positive effects compared to large firms. And there is a strong and precisely estimated effect on small firms.

To analyze firm concentration around the Primavera–Interlagos station, we constructed buffers of 250m, 500m, 1000m, and 2000m using the centroids of census tracts. This method enables spatial comparison by measuring the distance between tract centroids and the station.

Figure 9: Distribution of firms at different distances from the Primavera–Interlagos station.

Notes: The maps show the station location (blue dot), the centroids of census tracts (green dots), and the number of firms (shades of red). Dashed circles represent buffers of 250m, 500m, 1000m, and 2000m, respectively. The base map is provided by OpenStreetMap.

Table 7: Heterogeneity of Treatment Effects by Distance

Measure	**0-50m**	**0-100m**	**0-250m**	**0-500m**	**0-1000m**	**0-2000m**
ATT	46.5295***	15.4066**	12.9434**	6.9124*	1.9862	3.5415
	(0.9489)	(7.5283)	(5.5173)	(4.0133)	(5.5667)	(3.3551)
Treated Units	5	14	117	535	1679	3641
Groups	8	11	14	15	15	15
Observations	8,102	8,102	8,102	8,102	8,102	8,102
Control Mean	46.7511	46.7511	45.8558	44.4473	41.3617	39.4783

Notes: The ATT estimates represent the average treatment effect on the treated for different distance thresholds. Standard errors are in parentheses. The number of treated units refers to the census tracts that received treatment. The control mean represents the average number of establishments in the control group. * p<0.10, ** p<0.05, *** p<0.01.

Table 7 presents the heterogeneity of treatment effects by distance, showing the Average Treatment Effect on the Treated (ATT) at different distance thresholds. Here the distance definition is the distance between the census sector centroid and the nearest station. So if we consider 0-50m, we are considereing the effect on the census sector in wich centroids nearer than 50m of the closely station. The ATT estimates measure the impact of the treatment on the number of establishments within each specified distance from the intervention site.

The results indicate that the treatment effect is strongest at the shortest distance (0-50m), with an ATT of 46.53, significant at the 1% level. As the distance increases, the effect diminishes, with statistically significant impacts observed up to 500m. Beyond this distance, the ATT values decrease further and become statistically insignificant, suggesting that the treatment has a more localized effect.

The number of treated units (census tracts receiving treatment) increases with distance, as expected, with 5 treated units at 0-50m and 3,641 treated units at 0-2000m. The control mean, representing the average number of establishments in the control group, also decreases with distance, indicating that areas closer to the intervention site tend to have more establishments.

The results highlight the localized nature of the treatment effect, with stronger impacts near the intervention site and diminishing effects as the distance increases. The estimates suggest that economic activity is more responsive to the intervention

in closer proximity, whereas further areas experience negligible changes.

7 Mechanisms

The emergence of establishments following the creation of transportation stations can be attributed to several mechanisms. First, improved labor mobility enhances access to a broader workforce, facilitating business expansion. Second, transportation infrastructure expands consumer market reach, increasing demand for goods and services. Finally, the development of local economic ecosystems fosters business clustering, innovation, and synergies among firms, further stimulating economic activity.

7.1 Increase in Labor Mobility and Access to Workforce

With lower transportation costs for individuals, companies gain access to a more qualified and diverse workforce, as more workers can commute to the connected census tracts. This expansion in labor market reach enhances firm productivity and economic growth, as highlighted by (Combes et al., 2008), who emphasize the role of spatial wage disparities and worker sorting. Similarly, (Duranton and Puga, 2004) provide a micro-foundation for urban agglomeration economies, demonstrating how improved accessibility fosters firm clustering and economic dynamism.

The classification of labor-intensive and non-labor-intensive firms was based on a report from the World Bank that provides guidelines for categorizing sectors according to their labor intensity (World Bank, 2021). Labor-intensive firms, identified in sectors such as services and manufacturing that rely heavily on manual labor, require a higher ratio of workers to capital. Non-labor-intensive firms, on the other hand, are those that use less labor in relation to capital, typically found in industries such as technology or those that are more capital-intensive, where automation and machinery play a significant role. This classification allows for a differentiated analysis of how transportation infrastructure impacts sectors with varying labor requirements, based on the World Bank's framework

Table 8: Summary of Treatment Effects by Labor Intensity

Measure	**Intensive Labor**	**No Intensive Labor**
ATT	46.1706**	18.5043
	(19.6661)	(14.1898)
Treated Units	40	38
Groups	15	15
Observations	8,703	7,307
Control Group Mean	51.0023	22.2751

Notes: The ATT estimates represent the average treatment effect on the treated for labor-intensive and non-labor-intensive firms. Standard errors are in parentheses. The number of treated units refers to the census tracts that received treatment. The control group mean represents the average number of establishments in the control group. * p<0.10, ** p<0.05, *** p<0.01.

Table 8 presents the average treatment effect on the treated (ATT) for labor-intensive and non-labor-intensive firms. The ATT estimate for labor-intensive firms is 46.1706, with a standard error of 19.6661, indicating a statistically significant impact at the 5% level. For non-labor-intensive firms, the ATT is 18.5043, with a standard error of 14.1898, showing a positive but non-significative and smaller effect.

Labor-intensive sectors have more treated census tracts (40 vs. 38) and a larger sample (8,703 vs. 7,307). The control group mean is also higher for labor-intensive firms (51.00 vs. 22.28), indicating a stronger baseline presence. This suggests that labor-intensive firms tend to operate in areas with a higher baseline concentration of establishments.

Overall, the results indicate that the treatment has a stronger effect on labor-intensive firms, potentially due to their greater reliance on workforce accessibility and local agglomeration effects.

7.2 Expansion of Access to Consumer Markets

Improved transportation accessibility increases consumer flow to connected areas, stimulating the establishment of businesses in sectors such as retail, food services, and other local services. (Donaldson, 2018) highlights the transformative impact of transportation infrastructure on economic activity, demonstrating how improved connectivity fosters market integration and business expansion. Likewise, (Glaeser

and Gottlieb, 2009) emphasize the role of agglomeration economies in shaping urban prosperity, showing that enhanced accessibility contributes to spatial equilibrium and the concentration of economic activity in well-connected regions.

The businesses were classified into B2C (Business-to-Consumer) and B2B (Business-to-Business) categories based on the primary type of their clientele, using data from the company's main CNAE (National Classification of Economic Activities) code. B2C businesses, identified through CNAE codes associated with sectors like retail, food services, and personal services, directly serve individual consumers. In contrast, B2B businesses, classified under CNAE codes related to wholesale trade, manufacturing, and business services, primarily serve other businesses by providing goods or services as inputs for their operations. This classification allows for a more precise understanding of how transportation infrastructure impacts different types of economic activity.

Table 9: Summary of Treatment Effects

Measure	**B2C**	**B2B**
ATT	25.5311***	13.1056
	(9.2568)	(12.227)
Treated Units	40	37
Groups	15	15
Observations	8,509	6,465
Control Group Mean	36.91898	16.70228

Notes: The ATT estimate represents the average treatment effect on the treated. Standard errors are in parentheses. The number of treated units refers to the census tracts that received treatment. The control group mean represents the average value for the untreated units. * p<0.10, ** p<0.05, *** p<0.01.

The results in table 9 show that B2C firms experienced a significant increase of 25.53 businesses after station construction, while the effect on B2B firms (13.11 businesses) was smaller and not statistically significant. This suggests that improved transportation primarily benefits consumer-facing businesses such as retail and services. Additionally, the larger sample size for B2C (8,509 observations) compared to B2B (6,465) strengthens the reliability of the B2C estimates.

The control group had more B2C firms (36.92) than B2B firms (16.70) before

treatment, indicating that transportation investments reinforce existing economic patterns. The findings suggest that new stations boost businesses reliant on foot traffic, while B2B effects may take longer to emerge. The potential explanation is the effects was brought by people walkin since the station. These people is more propensity to demand B2C goods and services.

7.3 Development of Local Economic Linkages

More accessible public transportation facilitates interactions among entrepreneurs, workers, and consumers, fostering the emergence of economic clusters in connected areas. (Granovetter, 1985) emphasizes the role of social networks in economic action, highlighting how embedded relationships enhance business opportunities and innovation. Similarly, (Jacobs, 1969) argues that the dynamic exchange of ideas and resources in urban environments drives economic growth, reinforcing the importance of connectivity in stimulating local economic linkages.

Table 10: Summary of Treatment Effects

Measure	**HHI**
ATT	-0.1331**
	(0.0587)
Treated Units	6
Groups	15
Observations	126
Control Group Mean	0.2750

Notes: The ATT estimate represents summary of the estimated treatment effects on the Herfindahl-Hirschman Index (HHI), a measure of sectoral concentration of economic activities. Standard errors are in parentheses. The number of treated units refers to the census tracts that received treatment. The control group mean represents the average value for the untreated units. * p<0.10, ** p<0.05, *** p<0.01.

Table 10 shows that average treatment effect on the treated (ATT) is estimated at -0.1331, which is statistically significant at the 5% level (** p<0.05). The negative sign of the ATT indicates that the treatment led to a reduction in sectoral concentration, suggesting that economic activities became more diversified in the treated areas. The standard error of 0.0587 implies a relatively precise estimate.

The control group mean is 0.2750, indicating that, on average, control group areas exhibited a low level of sectoral concentration. The reduction in HHI due to treatment suggests that the intervention contributed to a more balanced distribution of economic activities across sectors, potentially fostering economic diversification.

To further investigate whether the construction of new stations fosters more complex and interconnected economic ecosystems, we construct a measure of functional diversity at the census tract level. This index captures the extent to which firms within a tract engage in economically complementary activities, based on sectoral input-output relationships. To operationalize this, we rely on Brazil's most recent national input-output matrix, published in 2015 by IBGE, which allows us to identify supply-chain linkages between CNAE (National Classification of Economic Activities) sectors.

Higher values of the index indicate a more functionally diverse business environment, meaning that firms in the area are not only from different sectors but also from sectors that tend to interact through production chains. This measure goes beyond sectoral variety by focusing on productive interdependence, thus enabling us to test whether public transportation infrastructure contributes to the formation of more integrated local economies.

Table 11: Summary of Treatment Effects

Measure	**Functional Diversity (I-O)**
ATT	-0.0064
	(0.0085)
Treated Units	40
Groups	15
Observations	8,703
Control Group Mean	0.1936

Notes: The ATT estimate summarizes the average treatment effect on the index of functional diversity based on productive complementarities (input-output related CNAEs). Standard errors are in parentheses. Treated units refer to census tracts that received treatment. The control group mean is the average index value among untreated units. * p<0.10, ** p<0.05, *** p<0.01.

Table 10 average treatment effect on the treated (ATT) is estimated at -0.0064, which is not statistically significant at conventional levels. The confidence interval

includes zero, indicating that there is no robust evidence that the treatment affected the formation of functionally diverse productive linkages. The standard error of 0.0085 suggests a relatively precise estimate, despite the null result.

The control group mean is 0.1936, reflecting a moderate level of functional diversity based on input-output complementarities between economic activities. The estimated ATT implies that, on average, treated areas did not experience a significant change in the structure of sectoral interdependence following the intervention. This suggests that fostering functional economic integration may require more than improvements in physical accessibility, such as additional institutional or economic incentives.

7.4 Population Density

One possible explanation for the impact of new transport infrastructure on urban dynamics is the intensification of population concentration around stations. The underlying logic is that improved accessibility may make certain areas more attractive for residential use, encouraging urban densification and potentially generating spillover effects on the viability of new business ventures.

This section investigates whether the introduction of transport stations is associated with increases in population density, exploring this mechanism as a channel through which urban space may be transformed.

Table 12: Treatment Effects on Population Density – Mechanism: Population Density

Measure	**Population Density**
ATT	0.0016***
	(0.0002)
Treated Units	40
Groups	15
Observations	8,703
Control Group Mean	0.0227

Notes: This table presents the average treatment effect on population density (inhabitants per square meter), as discussed in the section on the population density mechanism. The estimation uses the Callaway and Sant'Anna (2021) method with doubly robust estimators and not-yet-treated units as controls. The result is statistically significant at the 10% level. Standard errors are in parentheses. The control group mean corresponds to the average pre-treatment density for never-treated units. * p<0.10, ** p<0.05, *** p<0.01.

The results presented in Table 12 indicate that the introduction of transport stations is associated with a statistically significant increase in local population density. The estimated average treatment effect on the treated (ATT) is 0.0016, suggesting an average growth of 0.0016 inhabitants per square meter in treated census tracts following the intervention.

Given that the average population density among never-treated areas is 0.0227, this effect corresponds to an approximate 7% increase relative to the control group mean. The result is statistically significant at the 1% level, with a standard error of 0.0002, reinforcing the robustness of the estimated impact. These findings support the hypothesis that improved accessibility may lead to population densification around transit infrastructure.[6]

The observed increase in population density following station openings reinforces the other identified mechanisms. Denser areas tend to concentrate a larger number of consumers, amplifying the infrastructure's effect on B2C businesses. Moreover, a higher concentration of residents facilitates firms' access to the local labor force, enhancing the positive impact on labor-intensive sectors. Thus, population densification operates as a complementary channel that strengthens the effects of improved

[6]We need to mention that the population variable of CENSUS sector has interpolation, which may affect the effect of this mechanism.

mobility and consumer demand.

8 Robustness

To verify whether the results are sensitive to the composition of the control group, I re-estimated the model using only census tracts that were never exposed to transit station openings throughout the period as the control group. This excludes units that received treatment in later periods (not-yet-treated), thereby reducing potential contamination of the comparison group.

The estimated treatment effects remained robust in both magnitude and statistical significance. In some post-treatment periods, standard errors decreased slightly, which is consistent with a cleaner control group specification. This finding reinforces the internal validity of the results and suggests that future treatment does not bias the baseline estimates.

To ensure that the results were not driven by extreme values in specific areas, I re-estimated the model excluding census tracts located in Vila Olímpia, a region known for a surge in startups and business activity. The estimated treatment effects remained virtually unchanged, while standard errors decreased slightly. This suggests that the results are not sensitive to potential outliers and are representative of broader spatial patterns

To assess whether the estimated effects could have occurred by chance, a placebo test was performed in which the treatment years were randomly reassigned across treated units, while keeping the total number of station inaugurations per year constant. This simulation helps verify whether the observed effects could emerge by chance, as under this configuration, any detected impact cannot be attributed to the actual construction of transportation infrastructure. The results show no significant effects under this random assignment (ATT = 11.81 and SE = 13.45), suggesting that the original estimates capture a true causal relationship rather than statistical noise.

9 Conclusion

The overall average treatment effect (ATT) of a station results in an increase of 38.76 companies, representing a 78.46% increase compared to the control group. Additionally, the coefficient for the effect by time of exposure demonstrates robustness in the pre-treatment period and indicates a noticeable delay before the effect becomes positive following the station's construction. Specifically, the effect only becomes statistically significant from the fourth year onward. This delay is likely attributed to labor supply dynamics, as workforce adjustments in response to a new station may take time to materialize.

When analyzing the effect by year of adoption, the estimated impact was generally positive in years where significance was reached. However, for most adoption years, the effect was not statistically significant, suggesting heterogeneity in the station's impact depending on the timing of implementation. An outlier is Vila Olímpia station, which experienced a surge in technology startups at the turn of the century. The high standard error in certain years further indicates that stations built in the same year but in different locations may yield highly divergent effects. This highlights the importance of local context when evaluating policy impacts, particularly when fixing the adoption year and varying the census tract.

Expanding our perspective to heterogeneity analysis, the policy exhibits a positive impact across all sectors—commerce, industry, and services—with the most pronounced effect observed in the services sector. Regarding firm size, the policy primarily fosters the emergence of small and medium-sized enterprises, while larger firms remain largely unaffected. This suggests that service-oriented and smaller businesses are the most likely to proliferate around a newly established station (e.g., hair salons, maintenance services, computer shops, etc.).

Regarding the distance to the station, the closer the centroid of the census tract is to the station, the stronger the effect on the number of newly established companies. For census tracts located within 50 meters of a station, the effect is both substantial and highly significant. However, as the distance increases, the effect gradually weakens and loses statistical significance. Beyond 1 kilometer from the

station, no measurable impact on new business creation is observed.

Regarding mechanisms, we found that part of the effect is explained by labor supply in labor-intensive sectors, showing that the creation of a station is fundamental to the urban economy by facilitating worker commutes and enabling entrepreneurship in more areas. This is further supported by the heterogeneity in distance effects, where the impact is likely transmitted through individuals walking from the station. Additionally, the most affected sectors—services and small businesses—are generally highly labor-intensive.

When distinguishing between businesses targeting end consumers (B2C) and those focused on intermediate consumption (B2B), it becomes evident that part of the effect is also driven by demand generated at the station and transmitted to suppliers by foot. This finding is further reinforced by the heterogeneity in distance effects, which increases the equilibrium price and stimulates business creation.

Finally, a new station also fosters productive diversification in its vicinity, reinforcing the conclusion that its impact is significant and localized within a 1 km radius. This effect is driven by enhanced worker mobility, reduced labor costs, improved access to consumer demand, and increased productive diversification. Moreover, it is crucial to emphasize that small and medium-sized firms, particularly in the services sector, are the most likely to emerge following the construction of a new train or subway station.

References

Ahn, K., Jang, H., and Song, Y. (2020). Economic impacts of being close to subway networks: A case study of korean metropolitan areas. *Research in Transportation Economics*, 83:100900.

Albalate, D. and Bel, G. (2012). High-speed rail: Lessons for policy makers from experiences abroad. *Public Administration Review*, 72(3):336–349.

Allen, T. and Arkolakis, C. (2014). Trade and the topography of the spatial economy. *The Quarterly Journal of Economics*, 129(3):1085–1140.

Amior, M. and Manning, A. (2019). Commuting, migration and local joblessness. Discussion Paper CEP Discussion Paper No. 1645, Centre for Economic Performance, London School of Economics.

Atkin, D., Faber, B., and Gonzalez-Navarro, M. (2018). How do markets function? the role of intermediaries in a developing economy. *Econometrica*, 86(6):2003–2051.

Banerjee, S., De Vries, F. P., Hanley, N., and Van Soest, D. P. (2014). The impact of information provision on agglomeration bonus performance: an experimental study on local networks. *American Journal of Agricultural Economics*, 96(4):1009–1029.

Beaudry, P., Green, D. A., and Sand, B. M. (2018). In search of labor demand. *American Economic Review*, 108(9):2714–2757.

Callaway, B. and Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Combes, P.-P., Duranton, G., and Gobillon, L. (2008). Spatial wage disparities: Sorting matters. *Journal of Urban Economics*, 63(2):723–742.

Companhia do Metropolitano de São Paulo (2022). Relatório de expansão da linha 2-verde. Relatório técnico.

Companhia do Metropolitano de São Paulo (2023). Pesquisa origem e destino 2023. https://www.metro.sp.gov.br/pesquisa-od/. Accessed: June 21, 2025.

Companhia do Metropolitano de São Paulo (2025). Demanda – portal da transparência. https://transparencia.metrosp.com.br/dataset/demanda. Accessed: March 24, 2025.

Couto, C. (2011). Análise do projeto da linha 4 do metrô de são paulo. Trabalho de conclusão de curso, Universidade de São Paulo (USP), São Paulo, SP.

Desmet, K. and Rossi-Hansberg, E. (2014). Spatial development. *American Economic Review*, 104(4):1211–1243.

Donaldson, D. (2018). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5):899–934.

Donaldson, D. and Hornbeck, R. (2016). Railroads and american economic growth: A "market access" approach. *The Quarterly Journal of Economics*, 131(2):799–858.

Drucker, P. F. (1985). Entrepreneurial strategies. *California Management Review*, 27(2):9–25.

Duranton, G. and Puga, D. (2004). Micro-foundations of urban agglomeration economies. In *Handbook of Regional and Urban Economics*, volume 4, pages 2063–2117. Elsevier.

Duranton, G. and Puga, D. (2020). The economics of urban density. *Journal of Economic Perspectives*, 34(3):3–26.

Faggio, G. and Overman, H. (2014). The effect of public sector employment on local labour markets. *Journal of Urban Economics*, 79:91–107.

Felkner, J. S. and Townsend, R. M. (2011). The geographic concentration of enterprise in developing countries. *The Quarterly Journal of Economics*, 126(4):2005–2061.

Fenske, J., Kala, N., and Wei, J. (2023). Railways and cities in india. *Journal of Development Economics*, 161:103038.

Flyvbjerg, B. (2003). *Megaprojects and Risk: An Anatomy of Ambition*. Cambridge University Press, Cambridge.

Flyvbjerg, B. (2009). Survival of the unfittest: Why the worst infrastructure gets built—and what we can do about it. *Oxford Review of Economic Policy*, 25(3):344–367.

Flyvbjerg, B. and Gardner, D. (2023). *How Big Things Get Done*. Currency, New York.

Gaubert, C. (2018). Firm sorting and agglomeration. *American Economic Review*, 108(11):3117–3153.

Glaeser, E. L. and Gottlieb, J. D. (2009). The wealth of cities: Agglomeration economies and spatial equilibrium in the united states. *Journal of Economic Literature*, 47(4):983–1028.

Glaeser, E. L., Kerr, S. P., and Kerr, W. R. (2015). Entrepreneurship and urban growth: An empirical assessment with historical mines. *Review of Economics and Statistics*, 97(2):498–520.

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3):481–510.

Haddad, E. A., Hewings, G. J. D., Porsse, A. A., Van Leeuwen, E. S., and Vieira, R. S. (2015). The underground economy: Tracking the higher-order economic impacts of the são paulo subway system. *Transportation Research Part A: Policy and Practice*, 73:18–30.

International Association of Public Transport (2018). Metro benchmarking best practices. Technical report, UITP, Brussels. Accessed: 2025-06-21.

International Tunneling and Underground Space Association (2020). International tunneling and underground space association report. Technical report, International Tunneling and Underground Space Association (ITA). Accessed: June 21, 2025.

Jacobs, J. (1961). *The Death and Life of Great American Cities*. Random House, New York.

Jacobs, J. (1969). *The Economy of Cities*. Random House, New York.

Melo, P. C., Graham, D. J., and Noland, R. B. (2009). A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics*, 39(3):332–342.

Prud'Homme, R. and Bocarejo, J. P. (2005). The london congestion charge: A tentative economic appraisal. *Transport Policy*, 12(3):279–287.

Redding, S. J. and Sturm, D. M. (2008). The costs of remoteness: Evidence from german division and reunification. *American Economic Review*, 98(5):1766–1797.

Redding, S. J. and Turner, M. A. (2015). Transportation costs and the spatial organization of economic activity. In Duranton, G., Henderson, J. V., and Strange, W. C., editors, *Handbook of Regional and Urban Economics*, volume 5, pages 1339–1398. Elsevier.

Serviço Brasileiro de Apoio às Micro e Pequenas Empresas (SEBRAE) (2025). Data sebrae: Indicadores de empresas. `https://datasebraeindicadores.sebrae.com.br/resources/sites/data-sebrae/data-sebrae.html#/Empresas`. Accessed: March 24, 2025.

Shane, S., Locke, E. A., and Collins, C. J. (2003). Entrepreneurial motivation. *Human Resource Management Review*, 13(2):257–279.

Vuchic, V. R. (2005). *Urban Transit: Operations, Planning, and Economics*. John Wiley & Sons, Hoboken, NJ.

World Bank (2018). *Urban Rail Development Handbook*. World Bank Publications, Washington, DC. Accessed: June 21, 2025.

World Bank (2019). Infrastructure finance and public-private partnerships. Technical report, World Bank, Washington, DC. Accessed: June 21, 2025.

World Bank (2021). Labor content of exports (lacex) database. `https://databank.worldbank.org/source/labor-content-of-exports`. Accessed: 9 May 2025.